

Cluster Analysis of Total Assets Provided By Banks from Four Continents

Mirela Cătălina Türkeş

Faculty of Finance, Banking and Accountancy, Dimitrie Cantemir Christian University, Bucharest, Romania,

E-mail: mirela.turkes@ucdc.ro

Abstract The paper analyzed the total assets in 2016 achieved by the strongest 96 banks from 4 continents: Europe, America, Asia and Africa. It aims to evaluate the level of total assets provided by banks in 2016 and continental banking markets degree of differentiation to determine the overall conditions of the banks. Methodologies used in this study are based on cluster and descriptive analysis. Data set was built based on information reported by banks on total assets. The results indicate that most of total banking assets are found in Asia and the fewest in Africa. At the end of 2016, the top 16 global banks owned total assets of \$ 30.19 trillion according to the data set contains cluster 1 and the centroid was (2.25, 2.11, 3.06, 0.01).

Key words Cluster analysis, k-means clustering algorithms, banks, total banking assets, continents

JEL Codes: M31, G21

1. Introduction

In Europe, in order to enhance financial stability and reduce credit risks of all banks, they have adopted uniform regulations as increasing the share of total assets and capital. Practicing at the single monetary market, with low interest rates in 2016, helped to increase total asset quality of banks, stimulating the sustainability based on rising debt stock of specific borrowers in all sectors of the economies of European countries. Although the Asian banking system is totally different from that of the US and Europe, in 2016 it has substantially improved quality of bank total assets. Attracting funding sources which have been growing by small businesses, total asset growth resulted both banking institution and financial intermediaries (www.ey.com) In America, the evolution of banking net assets had a constant evolution in 2016, like 2015. Slow growth of bank assets was due to the increase of loans and securities assets (www.statistica.com). The banking sector in Africa is characterized by a high concentration. The first largest banks from most African countries hold the largest shares of total banking assets. Lack of competition is the main cause of the low level of profitability and volume of bank assets and capital (www.imf.org). This study has two objectives:

- Ranking of the 24 banks on every continent (Europe, America, Asia and Africa), according to the value of total assets in 2016, obtained three clusters (with large assets banks, with average assets banks and banks with assets less),
- Interpretation of defining characteristics of the data set analyzed; comparative evolution of total banking assets in 2016 of the top 24 banks on the continents. The objectives underlying the study took into account previous research conducted by many experts as: Sayaseng and Ercan (2016), Forte and Santos (2015), Vagizova *et al.* (2014), Knotek (2014), Park and Baik (2006) and other.

2. Literature review

Clustering involves using a set of techniques identifying subgroups in the data set. When we cluster the observations of a data set, we seek to partition them into distinct groups so that the observations within each group are quite similar to each other, while observations in different groups are quite different from each other (Gareth *et al.*, 2013). Clusters are two ways: K-means clustering and hierarchical clustering. In this study, it was chosen method K-means clustering. This is a simple technique of partitioning the data set in k distinct non-overlapping clusters. According to Mac Queen (1967) k-means clustering algorithms classify a lot of objects "x" in "k" prior number of clusters (MacQueen, 1967). K-means algorithm allows assigning each one of the observations exactly k clusters, depending on the similarities of data items. Cluster analysis was used by various distinguished researchers such as: Schiopu (2010), Forte and Santos (2015), Vagizova *et al.* (2014), Knotek (2014), Tudor *et al.* (2012), Zadeh *et al.* (2011) and Park and Baik (2006).

3. Methodology of research

This study aims banking market depth analysis of four different continents represented by four datasets (Europe, America, Asia and Africa). The study seeks to partition the data set into subsets (called clusters), given that certain data elements from a cluster may be similar to others in the same cluster and less similar data elements from the other clusters. In this case, the notion of "similar" may mean that banks are of the same banking group. To provide a degree of differentiation among banking markets, target opted for an analysis of similarity (cluster) for purposes of classifying data in Table 1 banking groups: large assets, medium assets and smaller assets processing the data through Excel into 3 clusters. Given

the type of data and purpose, it was chosen k-means clustering algorithm and it was used for the calculation of Euclidean distance.

Euclidean distance

The distance between two n-tuples in the data set will be calculated based on Euclidean metric. If $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ then the distance between x and y is defined by:

$$dist(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{1}$$

The distance is equivalent to minimizing the square of the distance, we will instead look at $dist^2(x, y) = (dist(x, y))^2$. If there are k clusters C_1, \dots, C_k with corresponding centroids c_1, \dots, c_k , then for each data element x of the k-means algorithm consists of finding the value j which minimizes $dist^2(x, c_j)$, [www.real-statistics.com/multivariate-statistics];

$$C_j = \{x: \min_n dist^2(x, C_n) = j\} \tag{2}$$

For define the new centroid c_j for cluster C_j we used the mean of all the elements in that cluster, so:

$$c_j = \frac{1}{m_j} \sum_{x \in C_j} x \tag{3}$$

Where m_j is the number of data elements in C_j .

Then tried to find centroids c_1, \dots, c_k , that minimize SS_E using formula:

$$SS_E = \sum_{j=1}^k \sum_{x \in C_j} dist^2(x, c_j) \tag{4}$$

Table 1. Degree of differentiation among banking markets

Nr	Top European Banks		Top American Banks		Top Banks in Asia		Top Banks in Africa	
	Bank	Total Assets (US\$b)	Bank	Total Assets (US\$b)	Bank	Total Assets (US\$b)	Bank	Total Assets (US\$b)
1	HSBC Holdings	2,608.15	JP Morgan Chase&Co	2,521.03	Industrial & Commercial B	3,549.88	Standard Bank Group	184.518
2	BNP Paribas	2,417.00	Bank of America	2,195.31	China Construction Bank	2,981.85	Absa Group	97.241
3	Deutsche Bank	2,006.71	Wells Fargo&Co	1,942.12	Mitsubishi UFJ Financial C	2,901.34	FirstRand	94.144
4	Credit Agricole Group	1,970.40	Citigroup Inc	1,818.12	Agricultural Bank of China	2,818.89	NedBank Group	80.110
5	Barclays PLC	1,819.61	Goldman Sachs Group	0.518	Bank of China	2,656.07	National Bank of Egypt	50.665
6	Societe Generale	1,624.97	Morgan Stanley	813.891	Japan Post Bank	2,022.02	Attirajiwafa Bank	40.026
7	Banco Snatander	1,494.39	U. S. Bancorp	454.134	Mizuho Financial Group	1,923.72	Banque Exterieur d'Alger	34.373
8	Groupe BPCE	1,357.34	Bank of New York Mellon	374.114	China Development Bank	1,904.34	Investec Bank	31.335
9	Royal Bank Scotland	1,214.11	PNC Financial Services	369.348	Sumitomo Mitsui Financial	1,764.69	Banque Misr	29.436
10	Lloyds Banking Group	1,142.22	Capital One Financial	345.061	Postal Savings Bank of Ch	1,203.37	Credit Populaire du Maroc	27.662
11	UBS Group AG	1,010.34	TD Group US Holding	338.720	Bank of Communications	1,200.63	Banque Marocaine du Cd	24.239
12	UniCredit SpA	992.04	HSBC North America Hol	304.439	Norinchukin Bank	999.623	Banque Nationale d'Alger	21.125
13	ING Group	985.57	State Street Corporation	256.140	Industrial Bank	861.606	Bank Muscat	18.774
14	Credit Suisse Group	838.55	BB&T Corporation	222.622	China Citic Bank	846.080	Libyan Arab Foreign Ban	18.000
15	BBVA	830.20	Charles Schwab	209.337	China Merchants Bank	835.592	Gumhouna Bank	17.513
16	Credit Mutuel	823.26	Suntrust Banks	205.091	Shanghai Pudong Devepol	819.541	First Bank of Nigeria	17.393
17	Intensa Sanpaolo	789.21	Ally Financial	157.397	China Minsheng Banking	792.262	Ecobank Group (Ecobank	17.162
18	Rabobank Group	764.04	American Express Compan	153.377	Agricultural Development	630.722	Al Baraka Banking Group	17.154
19	Nordea Bank	746.96	MUFG Americas Holding	151.117	Sumitomo Mitsui Trust Ho	580.308	Commercial International B	14.189
20	Standard Chartered Plc	660.99	Citizens Financial Group	147.015	China Everbright Bank	568.134	Zenith Bank	14.147
21	European Investment E	636.32	United Services Automobli	147.006	Resona Holdings	451.283	Credit Populaire d'Algerie	12.109
22	Commerzbank AG	592.68	Fifth Third Bancorp	143.279	State Bank of India	440.681	Banque de l'Agriculture et	12.010
23	DZ Bank	580.17	RBC SUA Holdco Corpora	142.593	Export-Import Bank of Ch	427.578	United Bank for Africa	11.901
24	KfW Group	569.98	Santanser Holdings USA	139.231	Ping An Bank Co	422.675	CNEP Banque (Caisse N	11.357

Source: www.statistica.com

In general, the K-means cluster analysis used 5 steps: (1) k number of clusters choice, (2) determining the initial selection of k centroids, (3) assign each data elements to its nearest centroid, (4) for each cluster determining a new selection of its centroid, (5) repeating the step3 until the centroids don't change.

4. Results and discussions

Then to perform cluster analysis data were classified into 3 clusters 4 columns. Then applied K Means clustering algorithm to the date of table 2 whit k = 3. The data includes 24 data elements (top 24 banks according to total net assets) that were presented in 4 dimensional points (four different continents - Europe, America Asia and Africa, see table 2). For the 3 clusters was done an initial assignment of the data elements. The result of the analysis was presented in Figure 2. It was presented in the first column as the cluster assignment for each of the 24 elements date. The centroid of the first cluster was (2.25, 2.11, 3.06, 0.11), the second cluster was (1.50, 0.40, 2.05, 0.03) and for the third cluster was (0.79, 0.20, 0.73, 0.01).

Top European Banks (EU)	Top American Banks (AM)	Top Banks in Asia (AS)	Top Banks in Africa (AF)	Cluster
2.61	2.52	3.55	0.18	1
2.42	2.20	2.98	0.10	2
2.01	1.94	2.90	0.09	3
1.97	1.82	2.82	0.08	1
1.82	0.00	2.66	0.05	2
1.62	0.81	2.02	0.04	3
1.49	0.45	1.92	0.03	1
1.36	0.37	1.90	0.03	2
1.21	0.37	1.76	0.03	3
1.14	0.35	1.20	0.03	1
1.01	0.34	1.20	0.02	2
0.99	0.30	1.00	0.02	3
0.99	0.26	0.86	0.02	1
0.84	0.22	0.85	0.02	2
0.83	0.21	0.84	0.02	3
0.82	0.21	0.82	0.02	1
0.79	0.16	0.79	0.02	2
0.76	0.15	0.63	0.02	3
0.75	0.15	0.58	0.01	1
0.66	0.15	0.57	0.01	2
0.64	0.15	0.45	0.01	3
0.59	0.14	0.44	0.01	1
0.58	0.14	0.43	0.01	2
0.57	0.14	0.42	0.01	3

Source: www.statistica.com

Figure 1. The assignment of data elements to the four continents and clusters

Cluster	Centroid	1	2	3	1	2	3	Dist-sq
1	EU	2.250565	1.502084	0.797501	0.531413	7.970838	16.57729	0.531413
2	AM	2.119145	0.402401	0.204162	0.040366	4.915797	11.62574	0.040366
3	AS	3.06299	2.054168	0.738672	0.117328	3.346329	9.165772	0.117328
4	AF	0.114003	0.037167	0.016982	0.229842	2.810224	8.311845	0.229842
5					4.843898	0.624801	4.763727	0.624801
6	k		3		3.184148	0.185467	2.703988	0.185467
7					4.648339	0.01976	1.952782	0.01976
8	Iter		10		5.192288	0.044233	1.701291	0.044233
9					5.828763	0.167879	1.253717	0.167879
10	SSE		3.445124		7.841444	0.856738	0.354742	0.354742
11					8.184514	0.974562	0.276864	0.276864
12					9.143153	1.382064	0.116014	0.116014
13					9.92617	1.710726	0.053186	0.053186
14					10.5145	1.932446	0.013563	0.013563
15					10.63542	1.97402	0.01049	0.01049
16					10.74319	2.024424	0.007204	0.007204
17					11.14961	2.161029	0.005128	0.005128
18					11.9993	2.633315	0.015352	0.015352
19					12.30765	2.806155	0.030456	0.030456
20					12.65033	2.98149	0.050993	0.050993
21					13.32653	3.384651	0.111864	0.111864
22					13.53953	3.498129	0.134481	0.134481
23					13.65282	3.563868	0.147831	0.147831
24					13.72623	3.600514	0.155868	0.155868

Figure 2. K-means cluster analysis

The squared distance of the first data elements to the first centroid was 0.53, lowest compared with the second cluster was 7.97 and the third cluster was 16.57. The variables included in columns 1, 2 and 3 are chosen elements of the lesser value (dist-sq), and then building new cluster. Cluster 1 is grouped 16 banks which hold the highest level of total net assets on four continents. The second cluster consists of 20 banks which rank second in size among total banking assets. The third cluster comprises the remaining of the 60 banks operating on four continents are ranked in third place in terms of total banking assets.

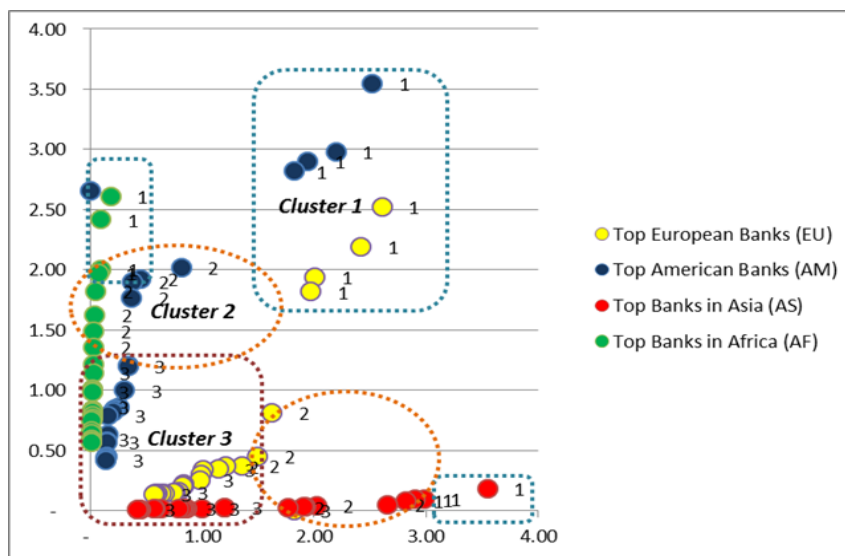


Figure 3. The assignment of data elements to the 3 clusters

Results of descriptive statistics indicate that the largest total assets of the first 24 banks are on the Asian continent. Within walking distance they are followed by banks in Europe and then America. Total record of banks' assets in Africa is \$0.89 trillion which is the lowest of those in 4 continents.

Table 4. General situation of the group of banks

SUMMARY

Groups	Count	Sum	Average	Variance	St. Deviation	Skewness	Kurtosis	Range	Minimum	Maximum
EU	24	28.4752	1.1865	0.3433	0.5986	1.0584	0.1431	2.0382	0.5700	2.6082
AM	24	13.5510	0.5646	0.5187	0.7357	1.8401	2.0018	2.5205	0.0005	2.5210
AS	24	33.6029	1.4001	0.8981	0.9681	0.8623	(0.5517)	3.1272	0.4227	3.5499
AF	24	0.8966	0.0374	0.0015	0.0402	2.5744	7.3778	0.1732	0.0114	0.1845

Skewness indications are positive, with values between 0.86 – 2.57 (trillion \$) which indicates a positive asymmetry. Kutosis indicator for banks in Asia is set to -0.55, indicating a flattened distribution compared to the normal sense. In Asia, the first banks in the system recorded total assets maximum of \$3.55 trillion and minimum of \$0.42 trillion. The first 24 organizations from top European banks, total assets recorded a maximum of 2.6 trillion \$ and a minimum of \$0.5 trillion, resulting a range of \$2.03 trillion. A large range observed in America where asset maximum total was \$2.52 trillion and minimum was \$0.5 billion. The small amplitude of total assets has been registered in Africa (\$0.17 trillion).

5. Conclusions

After realizing cluster and descriptive analysis, the following conclusions are resulted: (1) Cluster 1 included 16 banks which total assets of \$30.19 trillion, cluster 2 included 20 banks which total assets of \$19.98 trillion and cluster 3 recorded 60 banks which \$26.36 trillion; (2) First 24 banks in Asia hold total asset of \$33.6 trillion, 1.18 times more than in Europa and 2.47 times more than America; (3) Industrial & Commercial Bank of China (ICBC) holds total assets of \$3.54 trillion, the most powerful and the richest bank of 4 continents; (4) African banks registered in 2016 total assets \$896.6 billion, which is 15.22 times less than that in America; (5) The fewest assets are held by CNEP Banque (Caisse Nationale d' Epargne et de Prevoyance- Banque) from Africa.

References

- Forte, R. and Santos N. (2015). A cluster analysis of FDI in Latin America. *Latin American Journal of Economics*-formerly Cuadernos de Economía, 52 (1), pp. 25–56.
- Gareth, J., Witten, D., Hastie, T. and Tibshirani (2013). *An Introduction to Statistical Learning with Applications in R*, Springer Publisher, New York, pp. 360-362.
- Knotek, P. (2014). Banking Sectors in EMU – Cluster Analysis, *European Scientific Journal* December, 10(34), pp. 60-71.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, 281–297, University of California Press, Berkeley, Calif., <http://projecteuclid.org/euclid.bsm/1200512992>.
- Park, H. and Baik, K. (2006). A study for control of client value using cluster analysis, *Journal of Network and Computer Applications*, 29(4), pp. 262-276.
- Schiopu, D. (2010). Applying Two Step Cluster Analysis for Identifying Bank Customers' Profile, *Bulletin of University „Petrol – Gaze” from Ploiesti*, vol. LXII no. 3, pp. 66-75.
- Sayaseng, S. and Ercan H. (2016). The cluster analysis of the banking sector in Europe, Lengyel I. – Vas Zs. (eds) 2016: *Economics and Management of Global Value Chains*. University of Szeged, Doctoral School in Economics, Szeged, pp. 111–127.
- Tudor, A.I., Bara, A. and Andrei E. (2012). Clustering Analysis for Credit Default Probabilities in a Retail Bank Portfolio, *Database Systems Journal* vol. III, no. 2, pp. 23-30.
- Vagizova, V.I., Lurie, K.M., Ivasiv, I.B. (2014). Clustering of Russian banks: business models of interaction of the banking sector and the real economy. *Problems and Perspectives in Management* 12, pp. 83–93.
- Zadeh, R.B.K., Faraahi, A. and Mastali, A. (2011). Profiling bank customers behaviour using cluster analysis for Profitability, *Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management Kuala Lumpur, Malaysia*, January 22 – 24, 2011, pp. 458-467.

Web sources

- *** Descriptive statistics, available at https://en.wikipedia.org/wiki/Descriptive_statistics [accessed on 22 January 2017]
- *** Euclidian distance, available at http://en.wiktionary.org/wiki/Euclidean_distance [accessed on 22 January 2017]
- *** K – means Cluster Analysis, available at <http://www.real-statistics.com/multivariate-statistics/>, [accessed on 22 January 2017].
- *** Banking in Asia, available at [http://www.ey.com/Publication/vwLUAssets/EY-banking-in-asia-pacific/\\$FILE/EY-banking-in-asia-pacific.pdf](http://www.ey.com/Publication/vwLUAssets/EY-banking-in-asia-pacific/$FILE/EY-banking-in-asia-pacific.pdf), [accessed on 22 January 2017]
- *** Evolving banking trends in Africa, available at <https://www.imf.org/external/pubs/ft/dp/2015/afr1508.pdf>, [accessed on 22 January 2017]
- *** Largest banks in the United States in 2016, by assets (in billion U.S. dollars) available at <https://www.statista.com/statistics/250006/largest-banks-in-the-united-states-by-total-assets/>, [accessed on 22 January 2017]